



Universidad Nacional de Ingeniería (UNI)

Escuela Profesional de
Ciencia de la Computación
Sílabo 2027-I

1. CURSO

CS272. Bases de Datos II (Obligatorio)

2. INFORMACIÓN GENERAL

2.1 Curso	: CS272. Bases de Datos II
2.2 Semestre	: 5 ^{to} Semestre.
2.3 Créditos	: 3
2.4 Horas	: 1 HT; 4 HP;
2.5 Duración del periodo	: 16 semanas
2.6 Condición	: Obligatorio
2.7 Modalidad de aprendizaje	: Presencial
2.8 Prerrequisitos	: CS271. Bases de Datos I. (4 ^{to} Sem)

3. PROFESORES

Atención previa coordinación con el profesor

4. INTRODUCCIÓN AL CURSO

Este curso es una continuación de Gestión de Datos I, enfocándose en los mecanismos internos de los sistemas de gestión de bases de datos (DBMS) y tecnologías de bases de datos modernas. Cubre en profundidad la optimización de consultas, estructuras de indexación e integridad transaccional. Además, introduce a los estudiantes a modelos no relacionales (NoSQL) y arquitecturas distribuidas, que son esenciales para manejar datos a gran escala en entornos de nube contemporáneos.

5. OBJETIVOS

- Comprender las estructuras de almacenamiento interno y técnicas de indexación.
- Analizar y optimizar planes de ejecución de consultas.
- Comprender protocolos de transacción avanzados y mecanismos de recuperación.
- Evaluar e implementar soluciones usando modelos de bases de datos NoSQL.
- Comprender los principios de bases de datos distribuidas y paralelas.

6. RESULTADOS DEL ESTUDIANTE

AG-C08) Análisis de Problemas: Identifica, formula y analiza problemas complejos de computación. (Usage)

AG-C12) Usa enfoques sistémicos para seleccionar, desarrollar, aplica, integrar y administrar tecnologías seguras de computación para lograr los objetivos del usuario. (Usage)

7. TEMAS

Unidad 1: Procesamiento de Consultas (12 horas)	
Resultados esperados: AG-C08,AG-C12	
Temas	Aprendizaje esperado (<i>Learning Outcomes</i>)
▪ Estructuras de página ▪ Estructuras de índice: • árboles B+ • Índices hash: estáticos y dinámicos • Creación de índices en SQL ▪ Estructuras de archivos: • Archivos de montón (<i>heap</i>) • Archivos hash ▪ Algoritmos para operadores de consulta: • Ordenamiento externo • Selección • Proyección; con y sin eliminación de duplicados • Uniones naturales: Bucle anidado, Mezcla-ordenada, Unión hash • Análisis de eficiencia de algoritmos ▪ Transformaciones de consultas ▪ Optimización de consultas: • Rutas de acceso • Construcción de plan de consulta • Estimación de selectividad • Planes solo de índice ▪ Procesamiento de Consultas en Paralelo (por ejemplo, exploración paralela, unión paralela, agregación paralela) ▪ Afinamiento/rendimiento de base de datos: • Selección de índices • Impacto de índices en el rendimiento de consultas • Desnormalización	▪ Describir el propósito y organización de las estructuras de índice tanto de árbol B+ como hash [Describir] ▪ Componer un comando SQL para crear un índice (de cualquier tipo) [Componer] ▪ Especificar los pasos para los diversos algoritmos de operadores de consulta: ordenamiento externo, proyección con eliminación de duplicados, unión por mezcla-ordenada, unión hash, unión por bucle anidado por bloques [Analizar] ▪ Derivar el tiempo de ejecución (en solicitudes de E/S) para cada uno de los algoritmos anteriores [Analizar] ▪ Transformar una consulta en álgebra relacional a su equivalente apropiado para una ejecución canalizada de profundidad izquierda [Rediseñar] ▪ Calcular estimaciones de selectividad para una operación de selección y/o unión dada [Computar/Calcular] ▪ Describir cómo modificar una estructura de índice para facilitar una operación solo de índice para una relación dada [Describir] ▪ Para un escenario dado decidir qué índices soportar para la ejecución eficiente de un conjunto de consultas [Evaluar (valorar)] ▪ Describir cómo los DBMS aprovechan el paralelismo para acelerar el procesamiento de consultas dividiendo el trabajo entre múltiples procesadores o nodos [Describir]
Lecturas : [Silberschatz2019], [Ramakrishnan2003]	

Unidad 2: Internos de DBMS (12 horas)	
Resultados esperados: AG-C08,AG-C12	
Temas	Aprendizaje esperado (<i>Learning Outcomes</i>)
▪ Gestión de Búfer de DB ▪ Gestión de Transacciones: • Niveles de Aislamiento • ACID • Serializabilidad • Transacciones Distribuidas ▪ Control de Concurrencia: • Bloqueo en 2 Fases • Estrategias de manejo de bloqueos mutuos • Modelos de consistencia basados en quórum ▪ Gestor de Recuperación: • Relación con el Gestor de Búfer ▪ Control de Concurrencia: • Control de concurrencia optimista • Control de concurrencia por marca de tiempo ▪ Gestor de Recuperación: • Registro por escritura anticipada • Sistema de recuperación ARIES (Análisis, REDO, UNDO)	▪ Describir cómo un DBMS gestiona su Grupo de Búferes [Describir] ▪ Describir las cuatro propiedades para un gestor de transacciones correcto [Describir] ▪ Esbozar el principio de serializabilidad [Esquematizar/Esbozar]
Lecturas : [Silberschatz2019], [GarciaMolina2008]	

Unidad 3: Bases de Datos Distribuidas/Computación en la Nube (12 horas)	
Resultados esperados: AG-C08,AG-C12	
Temas	Aprendizaje esperado (<i>Learning Outcomes</i>)
■ DBMS Distribuido: ● Almacenamiento de datos distribuido ● Procesamiento de consultas distribuido ● Modelo de transacción distribuida ● Soluciones homogéneas y heterogéneas ● Bases de datos distribuidas cliente-servidor ■ DBMS Paralelo: ● Arquitecturas de DBMS paralelo: memoria compartida, disco compartido, nada compartido ● Aceleración y escalamiento, por ejemplo, uso del modelo de procesamiento MapReduce ● Replicación de datos y modelos de consistencia débil	■ Describir los componentes clave de un DBMS distribuido, incluyendo almacenamiento de datos distribuido, procesamiento de consultas y gestión de transacciones [Describir] ■ Analizar las ventajas y desventajas entre arquitecturas de DBMS paralelo: memoria compartida, disco compartido y nada compartido [Analizar] ■ Describir estrategias de replicación de datos y modelos de consistencia débil en sistemas de bases de datos distribuidas [Describir]
Lecturas : [Silberschatz2019], [Ozsu2020]	

Unidad 4: Sistemas NoSQL (12 horas)	
Resultados esperados: AG-C08,AG-C12	
Temas	Aprendizaje esperado (<i>Learning Outcomes</i>)
■ ¿Por qué NoSQL? (por ejemplo, Desajuste de impedancia entre Aplicación [CRUD] y RDBMS) ■ Modelo de datos Clave-Valor y Documento ■ Sistemas de almacenamiento (por ejemplo, sistemas Clave-Valor, Lagos de Datos (<i>Data Lakes</i>)) ■ Modelos de Distribución (Fragmentación y Replicación) ■ Bases de Datos de Grafos ■ Modelos de Consistencia (Actualización y Lectura, consistencia de quórum, teorema CAP) ■ Modelo de procesamiento (por ejemplo, Map-Reduce, map-reduce multi-etapa, map-reduce incremental) ■ Estudios de Caso: Sistema de almacenamiento en la nube (por ejemplo, S3); Bases de datos de grafos; "Cuándo no usar NoSQL"	■ Desarrollar un caso de uso para el uso de NoSQL sobre RDBMS [Crear] ■ Describir las características definitorias detrás de los modelos de datos basados en Clave-Valor y Documentos [Describir]
Lecturas : [Sadalage2012], [kleppmann2017designing]	

Unidad 5: Bases de Datos Semiestructuradas y No Estructuradas (12 horas)	
Resultados esperados: AG-C08,AG-C12	
Temas	Aprendizaje esperado (<i>Learning Outcomes</i>)
■ Datos no estructurados vectorizados (texto, video, audio, etc.) y almacenamiento vectorial: ● Vectorizador TF-IDF con n-grama ● Word2Vec ● Base de datos de arreglos o manejo de tipo de datos arreglo ■ Bases de datos semiestructuradas (por ejemplo, JSON): ● Almacenamiento: ○ Codificación y compresión de tipos de datos anidados ● Indexación: ○ árbol B, índice de salto, filtro Bloom ○ Índice invertido y compresión de mapa de bits ○ Indexación por curva de llenado de espacio para datos geo-semiestructurados ● Procesamiento de consultas para casos de uso OLTP y OLAP: ○ Compromisos de inserción, selección, actualización/eliminación ○ Estudios de caso sobre Postgres/JSON, MongoDB y Snowflake/JSON	■ Describir representaciones vectorizadas de datos no estructurados (por ejemplo, TF-IDF, Word2Vec) y sus modelos de almacenamiento [Describir] ■ Aplicar técnicas de indexación (por ejemplo, árbol B, índice invertido, filtro Bloom) a datos semiestructurados almacenados en formato JSON [Aplicar] ■ Analizar las ventajas y desventajas del procesamiento de consultas en bases de datos semiestructuradas para casos de uso OLTP y OLAP [Analizar]
Lecturas : [kleppmann2017designing], [manning2008introduction], [mikolov2013efficient], [agrawal2023toward], [mongodb-documentation]	

8. PLAN DE TRABAJO

8.1 Metodología

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

8.2 Sesiones Teóricas

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

8.3 Sesiones Prácticas

Las sesiones prácticas se llevan en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

9. SISTEMA DE EVALUACIÓN

Sesiones Teóricas:

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

Sesiones de Laboratorio:

Para verificar que los alumnos hayan alcanzado el logro planteado para cada una de las unidades de aprendizaje, realizarán actividades que les permita aplicar los conocimientos adquiridos durante las sesiones de teoría y se les propondrá retos que permitan evaluar el desempeño de los alumnos.

Exposiciones individuales o grupales:

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las

diferentes etapas de la evaluación del curso.

Lecturas:

A lo largo del curso se proporcionan diferentes lecturas, las cuales podrían ser evaluadas. El uso de herramientas Online permite a cada estudiante acceder a la información del curso, e interactuar fuera de aula con el profesor y con los otros estudiantes.

La nota final NF depende de de las siguientes notas:

- EP : Examen Parcial.
- EF : Examen Final.
- P : Evaluacion permanente (Promedio de prácticas y Trabajos calificados).

La nota final NF se calcula así:

$$NF = \frac{EP + EF + P}{3}$$

10. BIBLIOGRAFÍA BÁSICA